

Reconstruction of cysteine biosynthesis using engineered cysteine-free and methionine-free enzymes

Kendrick M Wang^{1,2,†}, Kosuke Fujishima^{3,4,5,†}, Nozomi Abe⁴, Kenji Nakahigashi⁴, Drew Endy¹,
Lynn J Rothschild^{6*}

¹Stanford University Department of Bioengineering, Stanford, CA, USA; ²Education Associates Program, Moffett Field, CA, USA; ³University Affiliated Research Center (UARC), NASA Ames Research Center, Moffett Field, CA, US; ⁴Institute for Advanced Biosciences, Keio University, Tsuruoka, Japan; ⁵Earth-Life Science Institute, Tokyo Institute of Technology, Tokyo, Japan; ⁶NASA Ames Research Center, Moffett Field, CA, USA; [†] These authors contributed equally to this work.

* Authors for correspondence: Lynn J. Rothschild

Planetary Science Branch, Mail Stop 239-20, NASA Ames Research Center, P.O. Box 1, Moffett Field, CA. 94035-1000, USA

Tel: +1 650 604 6525

E-mail: Lynn.J.Rothschild@nasa.gov

Alternate corresponding authors: Kosuke Fujishima or Kendrick Wang

Email: kosuke.fujishima@nasa.gov or kendrickwang@alumni.stanford.edu

Abstract

Ten of the proteinogenic amino acids can be generated abiotically while the remaining thirteen require biology for their synthesis. Paradoxically, the biosynthesis pathways observed in nature require enzymes that are made with the amino acids they produce. For example, *Escherichia coli* produces cysteine from serine via two enzymes that contain cysteine. Here, we substituted alternate amino acids for cysteine and also methionine, which is biosynthesized from cysteine, in serine acetyl transferase (CysE) and O-acetylserine sulfhydrylase (CysM). CysE function was rescued by cysteine-and-methionine-free enzymes and CysM function was rescued by cysteine-free enzymes. Structural modeling suggests that methionine stabilizes CysM and is present in the active site of CysM. Cysteine is not conserved among CysE and CysM protein orthologs, suggesting that cysteine is not functionally important for its own synthesis. Engineering biosynthetic enzymes that lack the amino acids being synthesized provides insights into the evolution of amino acid biosynthesis and pathways for bioengineering.

How life emerged from an inorganic prebiotic world to become the complex biological systems of today remains a pressing yet unsolved question. One problem within this grand puzzle is how the building blocks for polypeptides, which are the primary catalysts and structural elements for life, originated. Ten of the twenty standard proteinogenic amino acids—glycine, alanine, aspartic acid, glutamic acid, valine, serine, isoleucine, leucine, proline, and threonine—are found in multiple abiotic settings and thus are likely to have been present on prebiotic Earth^{1,2}. Some of these amino acids were transported to Earth via meteorites and comets, while others were generated via terrestrial geochemical events such as lightening, volcanism, and hydrothermal syntheses³.

While it has been hypothesized that short RNA fragments formed ribozymes to catalyze various chemical reactions⁴, there is no conclusive evidence for ribozymes involved in proto-metabolism where they would have synthesized the entire library of the modern amino acids, nor even if such syntheses pre- or post-dated the origin of RNA. The set of 20 amino acids is realized via chemical reactions encoded by protein enzymes, and in most cases these enzymes contain the amino acids they synthesize. However, it seems likely that in an earlier time polypeptides composed of a limited abiotically-derived set of amino acids might have served as biosynthetic enzymes. Such polypeptides could have played a role in developing and diversifying amino acid biosynthesis pathways.

Thus, we hypothesized that we could create biosynthetic enzymes that did not contain the amino acid they synthesized. Our aim was to construct enzymes that lack the amino acids they synthesize to demonstrate that it was possible to solve the “chicken-and-egg” puzzles found in modern amino acid biosynthesis. We focused on the biosynthesis of cysteine – an amino acid that plays an important role in modern protein structure and function. For example, two cysteine residues can form a disulfide bridge, a structure that results from the thiol group of cysteine and is unique among the twenty proteinogenic amino acids. Disulfide bridges are able to cross-link two cysteines in the same peptide, allowing for secondary structure, or join separate polypeptide chains forming tertiary structures⁵. Cysteine is also used to coordinate iron-sulfur cluster prosthetic groups that are found in key biochemical functions such as electron transfer, catalytic center and oxygen sensing⁶. In addition, cysteine is the key substrate providing the sulfur during

the biosynthesis of methionine. However, cysteine is not thought to arise via prebiotic synthesis nor is cysteine found in abiotic sources such as meteorites or cometary ice.

In plants and various bacterial and archaeal species, cysteine is synthesized via a two-step pathway from its precursor serine, which is derived from 3-phosphoglyceric acid in the glycolysis pathway (Figure 1). Although in some heterotrophic species of bacteria and eukaryotes cysteine is produced via the metabolism of methionine, methionine originally evolved from cysteine in autotrophs⁷. While the exact archaeal pathway remains unknown, genes similar to *E. coli* *cysE*, *cysK*, and *cysM* have been found in archaea⁸. In *E. coli* serine acetyltransferase (CysE) converts serine into O-acetylserine and then cysteine synthase A and B (CysK and CysM) incorporate sulfur from either hydrogen sulfide or thiosulfate to form cysteine. So far several attempts have been made to create functional proteins with simplified alphabets, such as a orotate phosphoribosyltransferase lacking seven amino acids (C, H, I, M, N, Q, and W)⁹ and in an extreme case a functional chorismate mutase using only nine amino acids¹⁰. We took a similar approach to design, synthesize, and validate the functionality of cysteine- and methionine-less *cysE*, and *cysM* genes using cysteine-dependent auxotrophic *E. coli* strains.

Results

Synthetic reconstruction of sulfur-free proteins involved in the cysteine synthesis pathway

To solve the “chicken-and-egg” biosynthetic puzzle for cysteine we designed a total of four modified synthetic genes *cysE-C*, *cysE-CM*, *cysM-C* and *cysM-CM*, corresponding to either cysteine-deficient (C) or cysteine-methionine (CM) deficient versions of *E. coli* *cysE* and *cysM* genes (Figure 2). Three and two cysteine residues in wild-type *E. coli* *cysE* and *cysM* genes were changed to serine, respectively. Serine was chosen due to its role as a precursor of cysteine (Figure 1) as well as its structural resemblance. Additionally, eight and eleven methionine residues were changed to leucine/isoleucine, starting from cysteine-deficient *cysE-C* and *cysM-C*. The methionine to leucine/isoleucine substitution is thought to be a ‘safe’ substitution that does not disturb protein structure and results in a similar hydrophobicity to methionine¹¹. Synthesized gene sequences were first PCR amplified using tagged primer sequences containing

HindIII and XhoI cut sites (Supplementary Figure 1). Each gene was then cloned into the pUC19 expression vector. Cloned gene sequences were further confirmed based on Sanger sequencing.

Structural analysis of CysE and CysM proteins

To explore the potential effects of these amino acid substitutions in the context of protein structure, we analyzed the crystal structures of the native CysE and CysM proteins. The quaternary structure of CysE (PDB: 1T3D) is a dimer of trimers, with each of the trimer subunits creating three separate active sites (Figure 3A). No cysteine-cysteine disulfide bridges are expected because the closest cysteine residues, C3 and C83, have a separation of 13.2 Å which is significantly greater than the accepted bridge length of 2.3 Å¹². In addition, several methionine-aromatic ring motif interactions, with a separation of ~5 Å, were observed that may stabilize protein structure¹³; one of such was between the M58 residue and the F131 residue (Figure 3A).

The quaternary structure of CysM (PDB: 2BHS) consists of two monomers that come together to form a dimer with two active sites (Figure 3B). Like CysE, CysM does not have any predicted cysteine-cysteine disulfide bridges. The closest two cysteine residues, C252 and C280, have a separation of 8.6 Å, which is greater than the 2.3 Å typical bridge length. CysM also has several methionine-aromatic ring motifs (e.g., M173 and F221) (Figure 3B).

By performing a search of related sequences and protein structures using HHblits we obtained multiple sequence alignments (MSA) for CysE and CysM. For CysE, a MSA with 1,918 sequences was obtained. This MSA had a probability of 100.0, E-value of 3×10^{-146} , P-value of 3×10^{-151} , score of 828.2, 0.0 SS, 273 Cols, 1-273 Query HMM, and 42-314 (314) template HMM. For CysM, a MSA with 124 sequences was obtained with a probability of 100.0, 1×10^{-160} E-value, 9×10^{-166} P-value, 907.8 score, 0.0 SS, 302 Cols, 1-303 Query HMM, and 4-315 (315) template HMM. The consensus amino acid for each residue in CysE and CysM was determined and displayed by HHblits (Supplementary Figure 2).

Auxotrophy of cysteine-dependent *E. coli* knockout strains

To determine if the cysteine biosynthesis enzymes lacking cysteine and methionine can enable biosynthesis of cysteine we obtained the K-12 *E. coli* single knockout strain Δ cysE from

the Keio collection¹⁴. We then constructed a double knockout strain, $\Delta cysK\Delta cysM$, by deleting the two homologous cysteine synthetase A and B genes, which are responsible for converting O-acetylserine to L-cysteine. We tested the auxotrophy of $\Delta cysE$ and $\Delta cysK\Delta cysM$ knockout strains by observing growth on LB medium, M9-glucose medium and M9-glucose medium supplemented with 0.5 mM L-cysteine. *E. coli* colonies were observed on LB and M9-glucose medium containing cysteine after 24 and 48 h respectively, while no growth was observed on M9-glucose minimal medium after 72 h. indicating that the knockouts strains are cysteine-dependent auxotrophs (Figure 4).

Cysteine-free proteins rescue cysteine-dependent knockout strains

Rescue experiments were conducted by transforming three *cysE* variants (*cysE*, *cysE-C* and *cysE-CM*) into the $\Delta cysE$ strain and three *cysM* variants (*cysM-opt*, *cysM-C* and *cysM-CM*) into the $\Delta cysK\Delta cysM$ strain. Wild-type *cysE* and codon optimized *cysM-opt* served as positive controls, whereas pUC19 vector with *lacZ* gene was used as a negative control. Transformants were plated on M9 glucose with ampicillin, kanamycin and IPTG supplemented and incubated at 30 °C. Colonies were observed on plates with the native *cysE* and *cysM* transformed cells and the synthetic *cysE-C*, *cysE-CM* and *cysM-C* transformed cells. These results suggest the recovery of *cysE* and *cysM* function using cysteine deficient enzymes. However, no growth was observed for plates with *cysM-CM* transformed cells. Negative controls, with *lacZ* transformed cells, also showed no growth (Figure 5). Initial transformation efficiencies for fully synthesized gene variants were low, possibly due to the errors during gene assembly or PCR amplification. Therefore, to confirm that the cysteine-dependent auxotrophs were rescued by the transformed synthetic genes, we isolated single colonies of *cysE-C*, *cysE-CM* and *cysM-C* from M9-glucose plates and expanded the colonies in liquid LB +Amp culture overnight at 37°C. The plasmids from these cells were then extracted for sequencing. Sequencing results indicated that all three genes recovered from rescued cells encoded the specific sequences as designed with no reverse mutations.. Likewise, we isolated *cysM-CM* transformed colonies from the LB +Amp plate (since no cells grew on M9-glucose) and confirmed that the plasmid insert sequence was exactly the as-designed *cysM-CM* sequence. The isolated plasmids were further re-transformed into new $\Delta cysE$ and $\Delta cysK\Delta cysM$ knockout cells and plated on M9-glucose +Amp +Kan +IPTG at 30°C along with the positive and negative controls. Colonies were again observed after 72 h indicating

that the observed rescue phenotype is specific to the synthetic genes encoded by the plasmids (Figure 5).

Growth curve analysis reveals shorter lag phase in cysteine-free gene transformants

To test whether the synthetic genes functioned better, worse, or the same as the modern genes that they had replaced, we performed comparative growth curve assays in liquid LB and M9 glucose media for four $\Delta cysE$ transformants, *cysE*, *cysE-C*, *cysE-CM*, and empty-pUC19 (Figure 6A and 6B), and four $\Delta cysK\Delta cysM$ transformants, *cysM-opt*, *cysM-C*, *cysM-CM*, and empty pUC19 (Figure 6C and 6D), in a 96-well plate for 48 h to observe how each synthetic gene affects distinct phases (lag, exponential, and stationary) during continuous culture. Cells were pre-cultured in LB +Amp and washed three times with 0.9% saline solution to remove all nutrients before inoculation to LB or M9 glucose medium +Amp + Kan supplemented with IPTG for gene expression. As predicted, all the transformants grew identically without a noticeable lag phase in the LB +IPTG condition. However, when grown in M9 glucose medium, auxotrophic strains were rescued by cysteine-free CysE-C and CysM-C have shown significantly shorter lag phase compared to native *E. coli* enzymes (Figure 6B and 6D).

Discussion

The focus of this study was to address the “chicken-and-egg” dilemma of cysteine biosynthesis by introducing synthetic cysteine-free and cysteine+methionine-free enzymes into modern *E. coli* deficient in these enzymes. The underlying hypothesis is that primordial polypeptides made from the abiotic subset of modern proteinogenic amino acids can serve as ancestral metabolic catalysts that, over evolutionary time scales, diversify and increase the repertoire of amino acids found today. Because no conclusive evidence for prebiotic cysteine and methionine has been reported, the most parsimonious explanation is that cysteine was created during the evolution of the metabolic network from its amino acid precursor serine, presumed to be a common proteinogenic amino acids in prebiotic settings³. Our rescue experiments using CysE-C and CysM-C enzymes imply that the acetylation of L-serine and the conversion of ester into cysteine can occur in a cysteine-independent manner. Interestingly, growth curve

experiments revealed a shorter lag phase of auxotrophs rescued by cysteine-free enzymes CysE-C and CysM-C compared to their wild-type analogs (Figure 6). Lag phase is a period which cells adjust their metabolic activity to the new environment and synthesize enzymes and factors necessary for cell division¹⁵. Since cysteine is no longer a limiting factor for CysE-C and CysM-C protein synthesis, rescued cells can carry out cysteine production immediately under cysteine-deprived condition. Thus, simplification of amino acid alphabets for the enzymes directly involved in the amino acid biosynthesis pathway may well be advantageous especially under conditions where a specific amino acid is absent.

We pursued reconstructing the same cysteine biosynthesis pathway using synthetic CysE and CysM enzymes lacking not only cysteine but methionine as well to extend our hypothesis that cysteine production can be carried out by not only cysteine-free enzymes but by entirely sulfur-free enzymes. Therefore, in addition to the previous cysteine substitution, we replaced all methionine residues with either leucine or isoleucine (Met to Leu/Ile), except the initiator N-Formylmethionine (fMet) (Figure 2). Accordingly, the CysE-CM protein, which consist of only 18 types of amino acids, has successfully complemented the loss of CysE function, whereas the CysE-CM protein failed to rescue the loss of CysE. Typically, methionine does not carry any catalytic activity due to its unbranched, hydrophobic side chain which contains a thioester; however, S/ π interactions between the sulfur atom and aromatic amino acids have recently been reported to increase the stability of a protein structure by 1-1.5 kcal/mol¹³. According to the CysM protein structure, two methionine-aromatic motifs were observed near the active center of the protein which could contribute to stabilizing the coils (173-176) and (217-222) as well as the α helix (177-187) and the β sheet (168-172) which are in close proximity and keeping the active center intact (Figure 3B). Furthermore, two other methionine residues, M95 and M116, are part of the active center known to interact with sulfate ions¹⁰. Therefore, replacement of these key methionine residues from CysM could have resulted in losing the ability to recognize and incorporation sulfate as well as inducing a structural alteration around the active site leading to loss of catalytic function or substrate recognition. In contrast, CysE does not have any cysteine or methionine residues within or near the active site; thus, the amino acid substitutions of cysteine and methionine should have less of an effect on the enzymatic function.

The ability to function without cysteine and methionine in CysE and without cysteine in CysM raises the question of the importance of these residues for enzymatic function in these enzymes. The consensus sequences of structurally aligned CysE orthologs and CysM orthologs, show that cysteine residues are not conserved at all, whereas some methionine residues are strictly conserved among the orthologous proteins (Supplementary Figure 2). Therefore we assume that during the evolution of the CysE and CysM cysteine was incorporated into the enzymes through random drift rather than providing benefits to their structural stability or function. However, at least two methionine residues, M95 and M119, are involved in the active center of CysM (Figure 3B) to maintain the structural integrity of active site and to interact with sulfur containing H₂S and thiosulfate¹⁶. The strict conservation among orthologous proteins suggests the involvement of methionine residues from early stage of CysM evolution (Supplementary Figure 2).

While it is becoming more feasible to create proteins with reduced alphabets¹⁷, there are no reports of proteins with a limited alphabet being directly involved in the biosynthesis of new amino acids. Thus, we believe this work to be the first example of enzymes that lack the amino acid they produce synthesizing that amino acid. While it is extremely unlikely that primordial cysteine synthesis enzymes were similar to the ones that we see now, the rational engineering of proteins via synthetic biology approaches have provided an instance for how biological pathways could have arisen from proteins utilizing a reduced set of amino acids. Similar studies should be explored for other amino acid biosynthesis pathways.

Materials and Methods

Gene design for cysteine deficient *cysE* and *cysM* genes

E. coli K-12 substrain MG1655 *cysE* (Genbank ID: 732686788) and *cysM* (Genbank ID: 732683705) genes were selected to create synthetic protein products without any cysteine. Gene sequences were modified using the software Geneious 8.0. Codons corresponding to CysE and CysM cysteine residues were substituted with serine codons in the gene sequence. To create the *cysE-C* gene, A total of three cysteine residues—located at amino acid residues 3, 23, and 83—were thus replaced with serine to create the *cysE-C* gene. Likewise, a total of two cysteine

residues—located at amino acid residues 252 and 280— were replaced with serine for *cysM-C*. We codon-optimized the *cysM-C* gene using the online codon optimization service provided by Integrated DNA Technologies (Coralville, IA USA) to achieve efficient expression in *E. coli*.

Design for cysteine-methionine deficient *cysE* and *cysM* genes

Enzymes lacking cysteine and methionine were created based on the *cysE-C* and *cysM-C* gene sequences. The codons for methionine residues were changed to those for either leucine or isoleucine. The *cysE-CM* gene was created by replacing the methionine codons by the leucine codon in all eight residues (26, 48, 58, 77, 155, 201, 254 and 256) of the *cysE-C* gene construct. Similarly, the *cysM-CM* gene was created by replacing the eight methionine codons of the *cysM-C* gene (residues 19, 48, 78, 103, 119, 173, 186, and 241) with leucine and three methionine codons with those for isoleucine (87, 95, and 129). *CysM-CM* was also codon-optimized using the online codon optimization service provided by Integrated DNA Technologies for efficient expression in *E. coli*.

Visualizing the structures of CysE and CysM proteins

Serine acetyltransferase ([EC 2.3.1.30](#); CysE), a 273 amino acid protein, is arranged as a dimer of loosely stacked trimers to carry out serine O-acetyltransferase function¹⁸. The cysteine synthase O-acetylserine sulfhydrylase ([EC 2.5.1.47](#); CysM), is a 303 amino acid protein exists as a dimer and is one of the two isozymes that catalyzes the second step of cysteine synthesis¹⁶. The crystal structures of CysE (PDB ID: 1T3D) and CysM (PDB ID: 2BHS) were obtained from the RSCB Protein Data Bank. PDB files were imported into PyMOL version 1.6.0.0 (Schrodinger, New York, NY). The residues involved with creating the active center pocket (catalytic core and substrate recognition) were identified. The active center residues are 92, 143, 157-158, 178, 184-185, 192, 204, 222, and 235 for CysE¹⁹ and 41, 68-72, 95, 119, 140, 141, 174, 175, 208-210, and 212 for CysM¹⁶. In addition to modeling, we performed a homology search by iterative HMM-HMM comparison using HHblits Release-2.18.2 (Tübingen, Germany) on the native *E. coli* CysE and CysM amino acid sequences. The multiple sequence alignment was performed with the following settings: FASTA alignment format, fraction of gaps <50%, uniprot20_Mar12 HMM database, 1 max iteration, global alignment mode, realign with MAC checked, and 0.0 MAC realignment threshold.

Gene synthesis, cloning, and transformation

Five different gene constructs were synthesized using the gBlocks Gene Fragments technology (Integrated DNA Technologies, Coralville, IA USA). The gene constructs are *cysE-C* (*cysE* with cysteine replaced), *cysE-CM* (*cysE* with cysteine and methionine replaced), *cysM-opt* (codon optimized *cysM* gene using the online codon optimization platform provided by IDT), *cysM-C* (codon optimized *cysM* with cysteine replaced), and *cysM-CM* (codon optimized *cysM* with cysteine and methionine replaced). For each gene construct, we added sequences containing HindIII and XhoI restriction sites at the 5' and 3' ends. PCR amplification of the gene construct was carried out with Q5 High-Fidelity Master Mix (New England Biolabs Inc., Ipswich, MA, USA) using the forward tag primer (5'- GCTTGCATCGTACGTATCGG-3'), and the reverse tag primer (5'- AGACGTAACGACCAACGCTAG-3'). Amplification was performed in a T100 Thermo Cycler (Bio-Rad Laboratories, Inc., Hercules, CA, USA) using a 30 s initial denaturation at 98 °C, 30 cycles of 10 s denaturation at 98 °C, 15 s annealing at 60 °C, and 30 s extension at 72 °C, and a final 2 min extension at 72 °C. PCR products were column purified using the GenCatch PCR Cleanup Kit (Epoch Life Science., Sugar Land, TX, USA) and verified using gel electrophoresis in a 1% (w/v) agarose gel in 1x TAE buffer run at 100V for 30 min. The resulting gel was stained with GelRed (Thermo Fisher Scientific Inc., Waltham, MA, USA).

For cloning and transformation, both the PCR products and the pUC19 vector (New England Biolabs) were digested using HindIII-HF and XhoI restriction enzymes at 37 °C for 1 h in 1x CutSmart buffer (New England Biolabs). The digested products were then cleansed of extraneous DNA using the MinElute Reaction Cleanup Kit (QIAGEN, Germantown, MD). Fifty nanograms of pUC19 and 50 ng of digested PCR products encoding the gene construct were mixed with ElectroLigase mix (New England Biolabs) in a final volume of 20 µl in 1x T4 DNA Ligase Reaction Buffer for 1 h. Ligated products were transformed into 50 µl of electrocompetent cells using an Electroporator 2510 (Eppendorf, Hauppauge, NY, USA) with an 1800V pulse. Cells were re-suspended in 500 µl of Super optimal broth with catabolite repression (SOC) medium (20 g tryptone, 0.5 g yeast extract, 0.5 g 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, and 10 mM MgSO₄ in 1L of deionized water adjusted to a final pH 7.0) and incubated for 1 h at 37 °C before plating.

Auxotrophic *E. coli* strains

We used two cysteine-dependent auxotrophic strains in this study. *ΔcysE* K-12 *E. coli*, a knockout strain, was provided through the in-house *E. coli* Keio Knockout Collection, which is a compilation of *E. coli* K-12 single-gene knockout mutants strains for all nonessential genes¹⁴. Since two O-acetylserine sulfhydrylases (A and B) are present in *E. coli* K-12 and unlike the description in the previous reports^{20, 21} that the single knockout strain for both genes (*cysK* and *cysM*) formed colonies on minimal medium without cysteine, we further constructed a *ΔcysKΔcysM* double knockout strain. First, the kanamycin marker was removed from the *ΔcysM* knockout strain, and then the *ΔcysM::Km* allele was introduced to the resulting strain through P1 transduction. Kanamycin resistant colonies were picked and the target strain, lacking both the *cysK* and *cysM* genes, was selected via PCR using primers specifically designed for *cysK* (JW2407_*cysK*-up; CCAGTATTGCGATTACCCC and JW2407_*cysK*-down; TCGATTTCAGCTTTGGCTTTT) and *cysM* (JW2414_*cysM*-up; CGAGCGTTTATTCGTTGGTC and JW2414_*cysM*-down; TTATCCGGCCTACAAAATCG). The *cysK cysM* double knockout strain did not form a colony on M9-glucose minimal medium for over 7 days. *CysK* and *cysM* are homologous genes that are responsible for the final step in cysteine biosynthesis; thus, *ΔcysKcysM* K-12 *E. coli* can serve as a cysteine-dependent conditional knockout strain.

Preparation of electrocompetent cells

A single colony each of *E. coli ΔcysE* and *ΔcysKΔcysM* mutants were selected from a fresh Luria Broth (LB) +Kanamycin (Kan) plate and inoculated in 10 mL of 2xYT medium (16 g casein digest, 10 g yeast extract and 5g NaCl per 1L of deionized water adjusted to final pH 7.0) and grown overnight at 37 °C as a starter culture. The overnight culture was inoculated into 1 L of 2xYT medium and incubated at 37 °C, shaking vigorously until it reached an OD₆₀₀ of 0.6. The culture was then put on ice and centrifuged at 4 °C at 2500 x g for 25 min. The cells were washed two times with 800 mL of 4 °C deionized water. The culture was then re-suspended in 80 mL of 10 % glycerol at 4 °C, and pelleted at 4000 rpm for 10 min at 4 °C. The supernatant was removed, and 2 ml of 10 % glycerol at 4 °C was added to the cells. For storage, the cells were dispensed into 100 μl aliquots, snap frozen in liquid N₂, and stored at -80 °C for future use.

Functional Screen for Cysteine Biosynthetic Enzymatic Function

The pUC19 vectors containing *cysE-C* or *cysE-CM* genes were transformed into electrocompetent $\Delta cysE$ cells, while vectors with *cysM-opt*, *cysM-C*, and *cysM-CM* genes were transformed into electrocompetent $\Delta cysK\Delta cysM$ cells. One microliter of DNA was added to 79 μ l of cells thawed on ice, and the mixture was allowed to incubate at 4 °C for 10 min. The cells were then electroporated at 1800 volts pulse using an Eppendorf Electroporator 2510 (Hauppauge, NY, USA), followed by incubation in 0.5 mL SOC medium at 37 °C for 15 min. After incubation, 0.5 μ l of 0.4 M isopropyl- β -D-thiogalactoside (IPTG) was added, and the culture was allowed to incubate for an additional 30 min. The cells were then pelleted at 7500 x g for 90 seconds, washed with 500 μ l of M9-glucose minimal medium, pelleted once again, and re-suspended in M9-glucose with 0.4 mM IPTG. The transformed cells were then plated on M9-glucose minimal medium supplemented with 0.4 mM IPTG, 50 μ g/ml kanamycin, and 100 μ g/ml ampicillin (M9-glucose +IPTG +Kan +Amp).

The plates were left to incubate at 30 °C for one week (168 h) or until colonies formed, whichever occurred first. Colonies that formed on minimal medium were isolated. Two controls were also conducted. A positive control consisted of electrocompetent $\Delta cysE$ cells transformed with *cysE* in pUC19 and the other with electrocompetent $\Delta cysK\Delta cysM$ cells transformed with codon optimized *cysM* in pUC19. The cells were allowed to incubate at 30 °C, overnight. The second positive control group supplemented the lack of cysteine through growing the knockout cells with empty pUC19 vectors on a fully supplemented media, LB +Amp, and were allowed to incubate at 37 °C overnight. Additionally, negative controls consisting of electrocompetent $\Delta cysE$ and $\Delta cysK\Delta cysM$ cells transformed with empty pUC19 expression vectors, were plated on minimal M9-glucose +IPTG +Kan +Amp media and allowed to incubate at 30°C for 168 h.

Colonies were picked and grown overnight in liquid LB +Kan +Amp media. Plasmids from rescued colonies were then purified using the QIAGEN QIAprep Spin Miniprep Kit. These plasmids were sequenced by Elim Biopharmaceuticals, Inc. (Hayward, CA US) with the forward primer 5'-CACTCATTAGGCACCCCAGG-3' and the reverse primer 5'-GAGACGGTCACAGCTTGTCT-3'. To confirm that the plasmids were responsible for rescuing the knock-out strains, the isolated plasmids and were re-transformed into respective new electrocompetent knockout cells. The cells were plated on M9-glucose +IPTG +Kan +Amp

plates and allowed to incubate at 30°C for 168 h or until colonies formed, whichever occurred first.

Growth curve analysis

A total of eight transformants: three *cysE* variants (*cysE*, *cysE-C* and *cysE-CM*) and empty pUC19 plasmid were transformed into the $\Delta cysE$ strain, and three *cysM* variants (*cysM-opt*, *cysM-C* and *cysM-CM*) and empty pUC19 plasmid were transformed into the $\Delta cysK\Delta cysM$ strain. Single colonies were isolated from each plate of transformants and grown to 0.4-0.6 OD₆₀₀ in LB +Amp +Kan at 37 °C. The cells were pelleted and washed with 0.9 % saline solution three times. Half of the cells were resuspended in LB +Amp +Kan and the other half of the cells were resuspended in M9-glucose +Amp +Kan. Excluding the transformants with empty pUC19 vector (controls), triplicate 200 µl cultures of each transformant in liquid LB +Amp +Kan +IPTG and M9-glucose +Amp +Kan +IPTG at 0.05 OD₆₀₀ were added into a 96-well plate respectively. Two 200 µl cultures of $\Delta cysE$ and $\Delta cysK\Delta cysM$ with empty pUC19 vectors and blanks of LB +Amp +Kan +IPTG and two 200 µl blanks of M9-glucose +Amp +Kan+ IPTG were also added to the 96-well plate. To minimize condensation, the lid of the 96-well plate was coated with Triton X-100 by adding 3-4 mL of 0.05% Triton X-100 in 20% ethanol to the lid and ensuring even coverage of the lid²². The solution was then poured off after 30 s, and the lid was allowed to air-dry against a vertical surface. The plate was covered and then placed into a SPECTRAMax Plus384 microplate reader (Molecular Devices, Sunnyvale, CA) and the OD₆₀₀ was taken for 48 hours with an interval of three minutes.

References

1. Danger G, Plasson R, Pascal R. Pathways for the formation and evolution of peptides in prebiotic environments. *Chem Soc Rev* **41**, 5416-5429 (2012).
2. Glavin DP, Callahan MP, Dworkin JP, Elsila JE. The effects of parent body processes on amino acids in carbonaceous chondrites. *Meteorit Planet Sci* **45**, 1948-1972 (2010).
3. Higgs PG, Pudritz RE. A thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code. *Astrobiology* **9**, 483-490 (2009).
4. Robertson MP, Joyce GF. The origins of the RNA world. *Cold Spring Harb Perspect Biol* **4**, (2012).
5. Sevier CS, Kaiser CA. Formation and transfer of disulphide bonds in living cells. *Nat Rev Mol Cell Bio* **3**, 836-847 (2002).
6. Beinert H, Holm RH, Munck E. Iron-sulfur clusters: nature's modular, multipurpose structures. *Science* **277**, 653-659 (1997).
7. Morzycka E, Paszewski A. Two pathways of cysteine biosynthesis in *Saccharomyces lipolytica*. *FEBS Lett* **101**, 97-100 (1979).
8. Borup B, Ferry JG. Cysteine biosynthesis in the Archaea: *Methanosarcina thermophila* utilizes O-acetylserine sulfhydrylase. *FEMS Microbiol Lett* **189**, 205-210 (2000).
9. Akanuma S, Kigawa T, Yokoyama S. Combinatorial mutagenesis to restrict amino acid usage in an enzyme to a reduced set. *Proc Natl Acad Sci U S A* **99**, 13549-13553 (2002).
10. Walter KU, Vamvaca K, Hilvert D. An active enzyme constructed from a 9-amino acid alphabet. *J Biol Chem* **280**, 37742-37746 (2005).
11. Lipscomb LA, *et al.* Context-dependent protein stabilization by methionine-to-leucine substitution shown in T4 lysozyme (vol 7, pg 772, 1998). *Protein Sci* **7**, 1481-1481 (1998).
12. Bhattacharyya R, Pal D, Chakrabarti P. Disulfide bonds, their stereospecific environment and conservation in protein structures. *Protein Eng Des Sel* **17**, 795-808 (2004).
13. Valley CC, *et al.* The methionine-aromatic motif plays a unique role in stabilizing protein structure. *J Biol Chem* **287**, 34979-34991 (2012).
14. Baba T, *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006 0008 (2006).

15. Rolfe MD, *et al.* Lag Phase Is a Distinct Growth Phase That Prepares Bacteria for Exponential Growth and Involves Transient Metal Accumulation. *J Bacteriol* **194**, 686-701 (2012).
16. Claus MT, Zocher GE, Maier THP, Schulz GE. Structure of the O-acetylserine sulfhydrylase isoenzyme CysM from *Escherichia coli*. *Biochemistry* **44**, 8620-8626 (2005).
17. Amikura K, Sakai Y, Asami S, Kiga D. Multiple Amino Acid-Excluded Genetic Codes for Protein Engineering Using Multiple Sets of tRNA Variants. *Acs Synth Biol* **3**, 140-144 (2014).
18. Hindson VJ, Moody PCE, Rowe AJ, Shaw WV. Serine acetyltransferase from *Escherichia coli* is a dimer of trimers. *J Biol Chem* **275**, 461-466 (2000).
19. Pye VE, Tingey AP, Robson RL, Moody PCE. The structure and mechanism of serine acetyltransferase from *Escherichia coli*. *J Biol Chem* **279**, 40729-40736 (2004).
20. Awano N, Wada M, Mori H, Nakamori S, Takagi H. Identification and functional analysis of *Escherichia coli* cysteine desulfhydrases. *Appl Environ Microbiol* **71**, 4149-4152 (2005).
21. Joyce AR, *et al.* Experimental and computational assessment of conditionally essential genes in *Escherichia coli*. *J Bacteriol* **188**, 8259-8271 (2006).
22. Brewster JD. A simple micro-growth assay for enumerating bacteria. *J Microbiol Methods* **53**, 77-86 (2003).

Acknowledgements

We are grateful to the NASA Ames Director's Discretionary Fund and the NASA Ames Science Innovation Fund for support. In addition, we would like to thank the Stanford University Undergraduate Advising and Research (UAR) Student Grants for funding as well as Professor Norman Sleep from the Department of Geophysics at Stanford University for his contributions. We also thank supports from the Tsuruoka City and Yamagata Prefecture to perform part of this study at Institute for Advanced Biosciences, Keio University.

Author contributions

K.M.W. performed overall biological experiments. K.F. designed and constructed the enzymes and wrote the manuscript. N.A. and K.N. constructed the double-knockout auxotrophic strain. K.F., D.A. and L.J.R. conceived and developed the idea. K.M.W., K.F., D.A. and L.J.R. wrote the manuscript. All authors contributed to provide comments on the results.

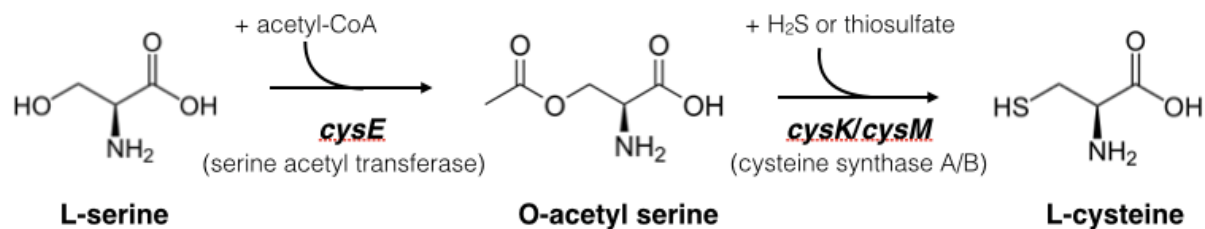
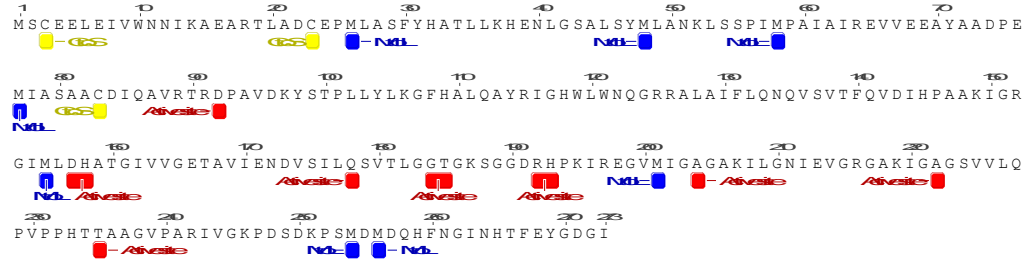


Figure 1. Two step cysteine biosynthesis pathway. The first step of cysteine synthesis is catalyzed by serine acetyltransferase, produced by the gene *cysE*. Serine acetyltransferase converts serine into O-acetylserine. From this step, O-acetylserine (thiol) sulphydrylase, produced by gene *cysK*, converts the intermediate into cysteine. O-acetylserine sulphydrylase function can also be provided by the homologous *cysM* gene. Thus, cysteine production in bacteria requires either *cysE* and *cysK* or *cysE* and *cysM*. In addition, genes similar to *cysE*, *cysK*, and *cysM* can be found within certain archaeal genomes (Kitabatake and So 2000).

A CysE protein



B CysM protein

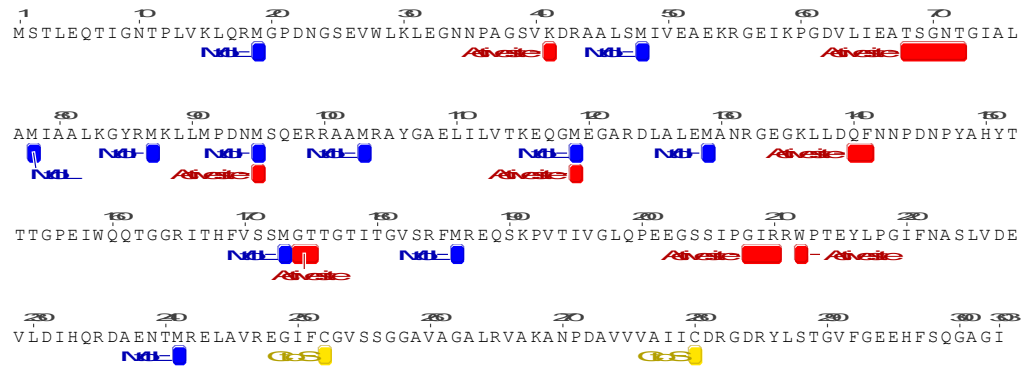


Figure 2. Cysteine and methionine residue substitution sites for *cysE* and *cysM* genes. A) Total of three cysteine residues (3, 23 and 83) and eight methionine residues (27, 48, 68, 77, 155, 201, 204 and 206) were replaced to serine and leucine, respectively. B) Total of two cysteine residues (252, and 280), 11 methionine residues (27, 48, 68, 77, 155, 201, 204 and 206) were replaced to serine and leucine/isoleucine, respectively. ‘C to S’ represents cysteine to serine substitutions (blue), ‘M to L’ represents methionine to leucine substitutions (red), ‘M to I’ represents methionine to isoleucine substitutions (orange), and ‘active center’ represents amino acids involved in substrate recognition and catalytic activity of the enzyme (black).

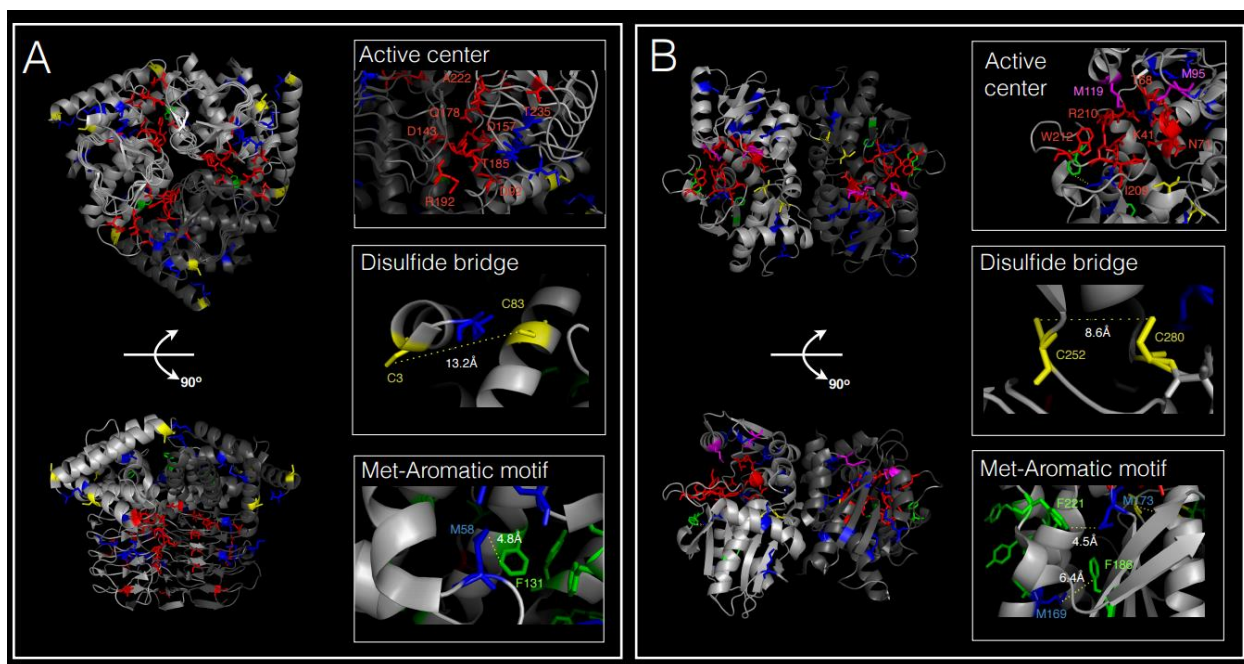


Figure 3. Key residues highlighted on the crystal structure model of CysE and CysM proteins. Cartoon diagram of the trimer A) *E. coli* acetyltransferase protein (CysE, PDB ID: 1T3D) and dimer B) *E. coli* cysteine synthetase B protein (CysM, PDB ID: 2BHS) are shown in gray scheme with each monomers in different exposure. The active site residues (red), substituted cysteine (yellow) and methionine (blue or purple) are highlighted in the left diagram. Methionines displayed in purple participate as an active site residue while methionines in blue are not part of the active site. Panels on the right represent the active center (top), the two closest cysteines (middle), and an example of methionine-aromatic motifs (bottom) found within the protein structures of CysE and CysM. Distances between each amino acid residue are denoted in angstroms.

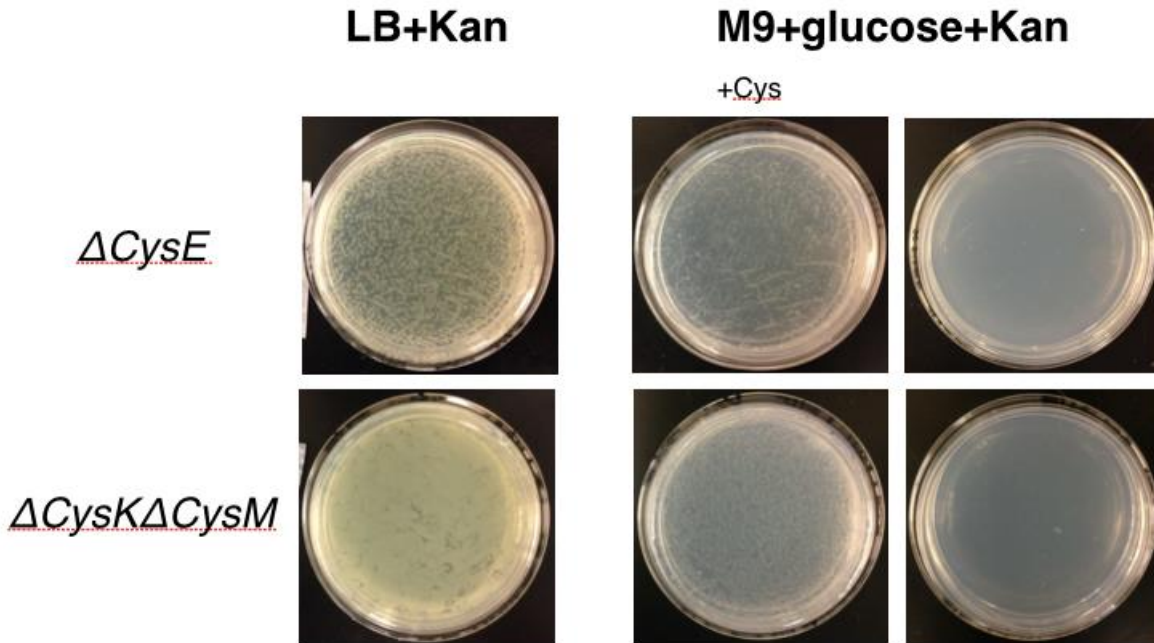


Figure 4. Cysteine-dependent growth of $\Delta cysE$ and $\Delta cysK\Delta cysM$ *E. coli* knockout strains.

Kanamycin resistant K-12 *E. coli* knockout strains $\Delta cysE$ and $\Delta cysK\Delta cysM$ were plated on LB medium supplemented with 50 $\mu\text{g/ml}$ kanamycin, M9-glucose medium supplemented with 0.5mM L-cysteine and 50 $\mu\text{g/ml}$ kanamycin, and M9-glucose medium with 50 $\mu\text{g/ml}$ kanamycin. For both strains, colonies were observed after incubation at 37°C for 24 h on LB +Kan and incubation at 30°C for 72 h on M9-glucose medium supplemented with L-cysteine (+cys) and kanamycin. Whereas, no growth was observed on M9-glucose medium supplemented with kanamycin.

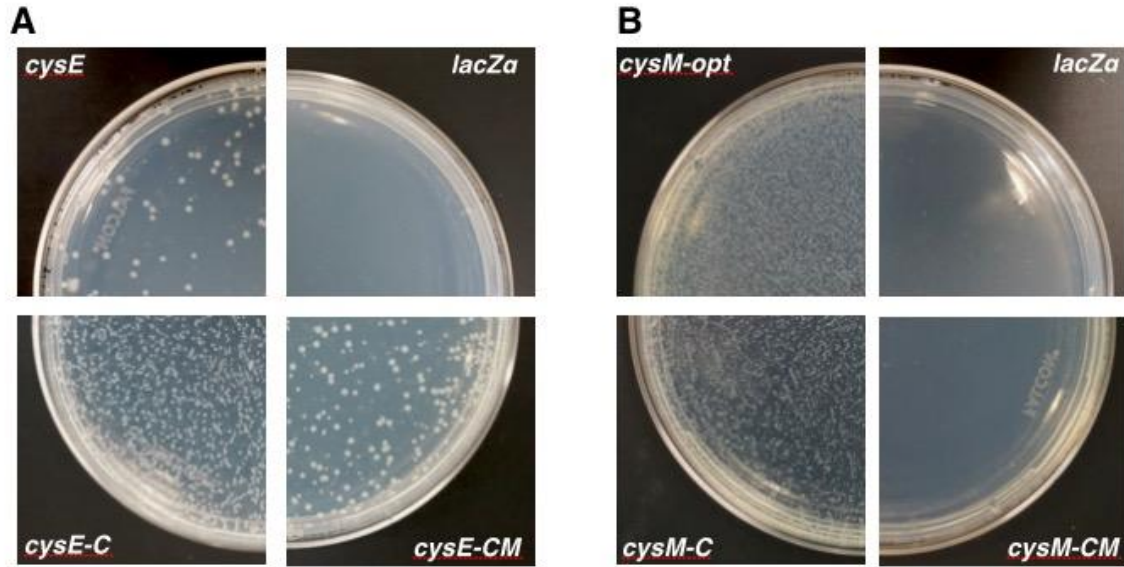


Figure 5. Synthetic *cysE* and *cysM* gene transformants display recovery of CysE function without cysteine and methionine and CysM function without cysteine. A) *E. coli* $\Delta cysE$ competent cells were transformed with positive control *cysE*, two *cysE* variants *cysE-C*/*cysE-CM* cloned into the multiple cloning site of pUC19 plasmid, and original pUC19 encoding N-terminal fragment of *lacZα* as a negative control. B) *E. coli* $\Delta cysM\Delta cysK$ competent cells transformed with positive control *cysM-opt*, two *cysM* variants *cysM-C*/*cysM-CM* in pUC19 plasmid, and original pUC19 encoding N-terminal fragment of *lacZα* as a negative control. Cells were plated on M9-glucose medium with 0.4 mM IPTG, 50 μ g/ml kanamycin, and 100 μ g/ml ampicillin and incubated at 30 °C for 72 h.

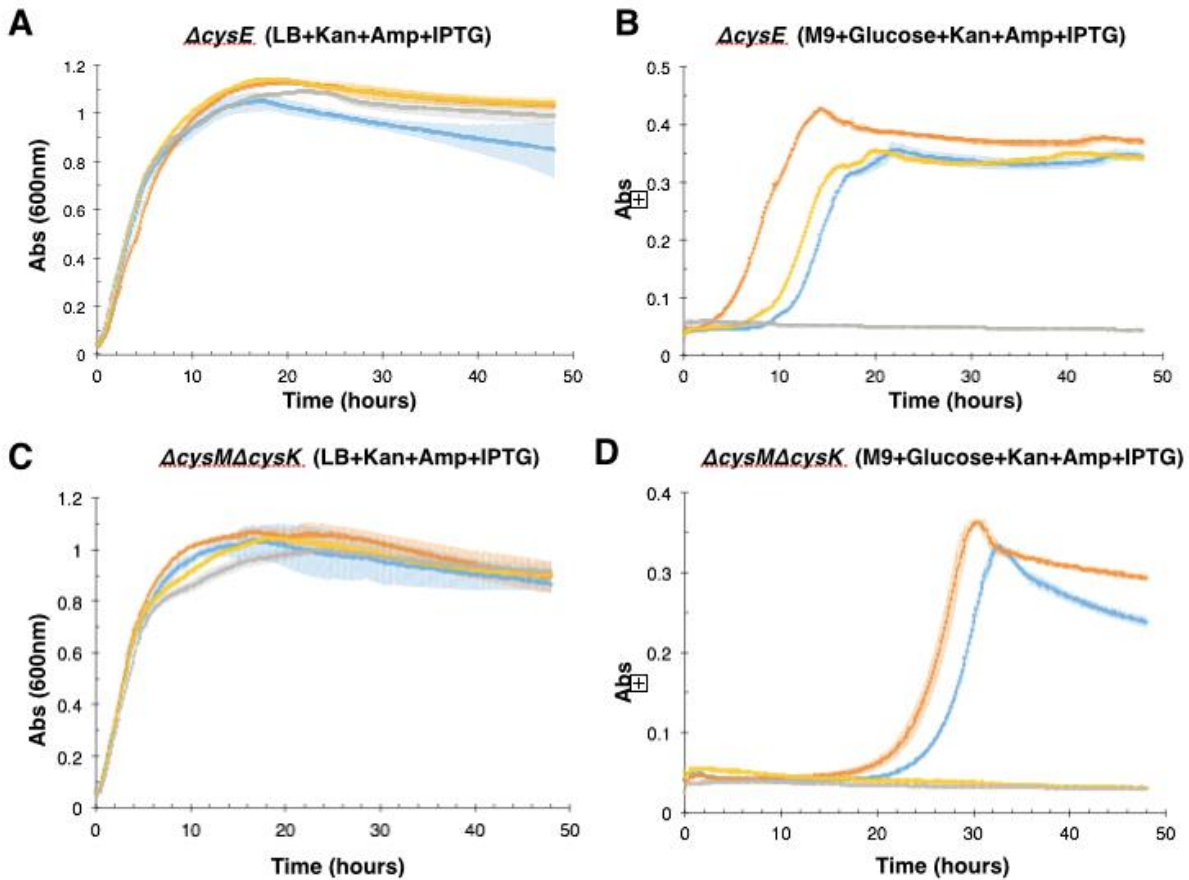


Figure 6. Growth curve of retransformed auxotrophic *E. coli* strains in LB and M9-glucose media. Each panel represents growth of cysteine-dependent *E. coli* auxotroph rescued by wild type enzymes: CysE and CysM (blue), cysteine-free enzymes: CysE-C and CysM-C (orange), sulfur-free enzymes: CysE-CM and CysM (yellow), and LacZ α protein expressed from pUC19 plasmid (gray). Growth curve was monitored at absorbance 600 nm using 96-well plate. Standard deviations of the growth curves are displayed and are calculated from triplicates.